INTRUSION DETECTION SYSTEM BY USING FC-ANN

MR SAGAR R DESHMUKH, PROF PAVAN N MUNDHARE
Dept. of Comp Network Engg (CNE), KSIET hingoli, SRT MUN University

Accepted Date: 05/03/2015; Published Date: 01/05/2015

Abstract: An intrusion detection system (IDS) is a software application that monitors network or system activities for malicious activities. The research on neural network methods and machine learning techniques to improve the network security by examining the behavior of the network as well as that of threats is done in the rapid force. There are several techniques for intrusion detection which exist at present to provide more security to the network, however many of those are static. Many researchers used machine-learning techniques for intrusion detection, but some shows poor detection, some techniques takes large amount of training time. In this paper, we study a learning approaches i.e. neural network approaches used for intrusion detection in the recent research papers has been surveyed and proposed an extreme learning approach to solve the training time issue.

Keywords: ANN, IDS, HIDS, NIDS
INTRODUCTION

Intrusion Detection System is any hardware, software, or a combination of both that monitors a system or network of systems against any malicious activity. This is mainly used for detecting break-ins or misuse of the network. In short, we can say that IDS is the ‘burglar alarm’ for the network because much like a burglar alarm, IDS detects the presence of an attack in the network and raises an alert. An IDS provides three functions: monitoring, detecting and generating an alert. IDS are often considered as the functionality of firewall. But there is a thin line of difference between them. A firewall must be regarded as a fence that protects the information flow and prevent intrusions whereas IDS detects if the network is under attack or if the security enforced by the firewall has been breached. Together firewall and IDS enhance the security of network. Intrusion Detection System uses a security policy (or rules) to detect unusual activity. These rules are defined by the administrator based on the needs of the organization. Any activity that violates this security policy will be considered a security threat and will be reported to the administrator via email or as page or as SNMP traps. These policies must be updated regularly to keep up with the threats and needs. Of the security incidents that occur on a network, the vast majority (up to 85 percent by many estimates) come from inside the network. These attacks may consist of otherwise authorized users who are disgruntled employees. The remainder come from the outside, in the form of denial of service attacks or attempts to penetrate a network infrastructure. Intrusion detection systems remain the only proactive means of detecting and responding to threats that stem from both inside and outside a corporate network. Intrusion detection is a major focus of research in the security of computer systems and networking. An intrusion detection system (ids) [1] is used to detect unauthorized intrusions i.e. attacks into computer systems and networks. These systems are known to generate alarms (alerts).the following general terms used for detection and identification of attack and non-attack behavior.

- True Positive (tp): the amount of attack detected when it is actually attack;
- True Negative (tn): the amount of normal detected when it is actually normal;
- False Positive (fp):the amount of attack detected when it is actually normal called as false alarm;
- False Negative (fn): the amount of normal detected when it is actually attack, namely the attacks which can be detected by intrusion detection system.
Artificial Neural Network (ANN) is one of the widely used techniques and has been successful in solving many complex practical problems and ANN has been successfully applied into IDS [2]. However, the main drawbacks of ANN-based IDS exist in two aspects: (1) lower detection precision, especially for low-frequent attacks, e.g., Remote to Local (R2L), User to Root (U2R), and (2) weaker detection stability.

For the above two aspects, the main reason is that the distribution of different types of attacks is imbalanced. For low-frequent attacks, the leaning sample size is too small compared to high-frequent attacks. It makes ANN not easy to learn the characters of these attacks and therefore detection precision is much lower. In practice, low frequent attacks do not mean they are unimportant. Although prior research has proposed some approaches, when encountering large datasets, these approaches become not effective.

To solve the above two problems, we propose a novel approach for ANN-based IDS, FC-ANN, to enhance the detection precision for low-frequent attacks and detection stability. The general procedure of FC-ANN approach has the following three stages. In the first stage, a fuzzy clustering technique is used to generate different training subsets. Based on different training sets, different ANNs are trained in the second stage. In the third stage, in order to eliminate the errors of different ANNs, a meta-learner, fuzzy aggregation module, is introduced to learn again and combine the different ANN’s results. The whole approach reflects the famous philosophy “divide and conquer”. By fuzzy clustering, the whole training set is divided into subsets which have less number and lower complexity. Thus the ANN can learn each subset more quickly, robustly and precisely, especially for low-frequent attacks, such as U2R and R2L attacks.

I. Literature Survey

In paper Vladimir Bukhtoyarov a proposed a neural network ensemble approach to detect intrusion. The approach is used for fixed-size neural networks ensembles with single stage voting. To overcome the problem of detecting the network attacks collective neural network approach is used. But the structure become complex due to collective approach and more amount of training time requires for training each ANN model which are issues of the system. The choice of the threshold to appeal to the neural network ensemble classifier is one of the issues[1].

Prof. D.P. Gaikwad, an FC-ANN approach in based on ANN and fuzzy clustering to solve the lower detection precision, weaker detection stability issues. In the proposed model restore point is provided for rolling back of system files, registry keys, installed programs and the project data base etc. To reduce the complexity and size of the subsets, first different training
subsets are generated by using fuzzy clustering. Then for those subsets different ANN models are trained and finally results are combined [2].

V. Jaiganesh, proposed a back-propagation approach to detect intrusion in. First the input and its corresponding target are called a Training Pair is generated. Then the training pair is applied to the network. Detection rate and false alarm rate are the performance measure used for evaluation of proposed method. The detection rate for DoS, Probe, U2R, R2L attack is below 80%. Poor detection of attackers if some hidden attackers are present is one of the issues [3].

FC-ANN [4] is hierarchical IDS based neural network and fuzzy clustering. It is composed of three layers. The first layer is a fuzzy clustering that generates the different training subsets. The second layer represents the different neural networks that are trained to formulate different base models. The last layer is a fuzzy aggregation module, which is employed to aggregate these results and reduce the detected errors.

Types of IDS

Host Based IDS

Intrusion Detection System is installed on a host in the network. HIDS collects and analyzes the traffic that is originated or is intended to that host. HIDS leverages their privileged access to monitor specific components of a host that are not readily accessible to other systems.

Network Based IDS

Network IDSs (NIDS) are placed in key areas of network infrastructure and monitors the traffic as it flows to other host. Unlike HIDS, NIDS have the capability of monitoring the network and detecting the malicious activities intended for that network. Monitoring criteria for a specific host in the network can be increased or decreased with relative ease. NIDS should be capable of standing against large amount number of network traffic to remain effective. As network traffic increases exponentially NIDS must grab all the traffic and analyze in a timely manner.

Stack Based IDS

Stack based IDS is latest technology, which works by integrating closely with the TCP/IP stack, allowing packets to be watched as they traverse their way up the OSI layers. Watching the packet in this way allows the IDS to pull the packet from the stack before the OS or application has a chance to process the packets.
Signature-Based IDS

Signature-Based IDS use a rule set to identify intrusions by watching for patterns of events specific to known and documented attacks. It is typically connected to a large database which houses attack signatures. It compares the information it gathers against those attack signatures to detect a match.

Anomaly Based IDS

Anomaly-Based IDS examines ongoing traffic, activity, transactions and behavior in order to identify intrusions by detecting anomalies. It works on the notion that “attack behavior” differs enough from “normal user behavior” such that it can be detected by cataloging and identifying the differences involved. In most anomaly-based IDS’s the system administrator defines the baseline of normal behavior. This includes the state of the network's traffic load, breakdown, protocol, and typical packet size.

Attacks go down into four categories:

(1) Denial of Service (DoS): making some computing or memory resources too busy to agree to legitimate users access these resources.

(2) Probe (PRB): host and port scans to get together information or get known vulnerabilities.

(3) Remote to Local (R2L): unauthorized access from a remote machine in order to utilize machine’s vulnerabilities.

(4) User to Root (U2R): unauthorized access to local super user (root) privileges using system’s vulnerability.

GOAL

Our Goal is to build a high performance IDS that better detects the low-frequent attacks without losing their high performance on the detection of frequent attacks and normal behavior. Unlike some related works, the built model must give a high performance and train in a very short time.

PROBLEM STATEMENT

Providing security in a Distributed System requires more than user authentication with passwords or digital certificates and confidentiality in data transmission. Distributed model of cloud makes it vulnerable and prone to sophisticated distributed intrusion attacks like
Distributed Denial of Service (DDoS) and probe. To handle large scale network access traffic and administrative control of data and application in cloud, a new multi-threaded distributed cloud IDS model has been proposed. Our proposed cloud IDS handles large flow of data packets, analyze them and generate reports efficiently by integrating knowledge and behavior analysis to detect intrusions. After detecting that intrusions we have check his authentication by taking Turing Test and CAPTCHA work and if not authenticated then block him.

**Existing System**

An intrusion is defined as any set of actions that attempt to compromise the integrity, confidentiality, or availability of a resource. An Intrusion Detection System (IDS) monitors and restricts user access to the computer system by applying certain rules. These rules are based on expert knowledge extracted from skilled administrators, who construct attack scenarios and apply them to find system exploits. The system identifies all intrusions by users and takes or recommends necessary action to stop an attack on the database. Two approaches to intrusion detection are currently used. The first one, called misuse detection, is based on attack signatures, i.e., on a detailed description of the sequence of actions performed by the attacker. This approach allows the detection of intrusions matching perfectly the signatures, so that new attacks performed by slight modification of known attacks cannot be detected. The second approach is based on statistical knowledge about the normal activity of the computer system, i.e., a statistical profile of what constitutes the legitimate traffic in the network. In this case, intrusions correspond to anomalous network activity, i.e. to traffic whose statistical profile deviates significantly from the normal one.

**PROPOSED SYSTEM**

To overcome this drawback we are developing new model of Intrusion Detection System which has capacity of self detecting or updating attacks. In proposed IDS model we are develop Artificial Neural Network algorithm with fuzzy logic to detect and update database for newly attacks. in proposed model we define two separate set of data. 1] Training set 2] Testing set. In training set every user query checked using apriori algorithm and fuzzy algorithm .In training set we use apriori, artificial neural network, clustering algorithm for train the user query and database.

**Proposed Methodology**

The proposed approach has the following three phases.

1) Data pre-processing: Convert raw data to machine readable form.
2) Training: In this phase, the network will be trained on normal and attack data.

3) Testing: Activity will be predicting i.e. either intrusive or not.

Fig.1. Proposed Architecture of IDS.

The architecture has following modules.

Network Data Monitoring:

This module will monitor network stream and capture packets to serve for the data source of the NIDS.

Pre-processing:

In pre-processing phase, network traffic will be collected and processed for use as input to the system.

Feature Extraction:

This module will extract feature vector from the network packets (connection records) and will submit the feature vector to the classifier module. The feature extraction process consists of feature construction and feature selection. The quality of feature construction and feature selection algorithms is one of the most important factors that influence the effectiveness of IDS. Achieving reduction of the number of relevant traffic features without negative impact on classification accuracy is a goal that largely improves the overall effectiveness of the IDS.
Classifier:

This module will analyze the network stream and will draw a conclusion whether intrusion happens or not. BPN and ELM techniques can be used as a classifier. The most successful application of neural network is classification or categorization and pattern recognition.

Training:

The learning process is the process of optimization in which the parameters of the best set of connection coefficients (weights) for solving a problem are found.

Testing:

When detecting that intrusion happens, this module will send a warning message to the user.

Knowledgebase:

This module will serve for the training samples of the classifier phase. The Artificial Neural Networks can work effectively only when it has been trained correctly and sufficiently.

PROPOSED ARCHITECTURE

Fig 2 PROPOSED IDS Architecture
In our system we are dividing our system in different modules which are listed below,

1) Checking source
2) Counting
3) Attack detection
   i) FC Module (Fuzzy Clustering Module)
   ii) Ann module
   iii) Fuzzy Aggregation Module
4) Turing Test Module
5) Question Generation Module

**Brief Description of each module**

1) **Checking source**

   In this module we are checking the source of attack. We are providing authentication for client for login. If client attacks with some pattern then by identifying that client's IP address we finding its source.

2) **Counting**

   In this module we are recording the source address destination address and the time at which client performs login test. After login successful the counting module is reset. It will be enable by the Attack Detection module when there are some suspected traffic been detected.

3) **Attack detection**

   In this section, we elaborate our new approach; FC-ANN. FC-ANN firstly divides the training data into several subsets using fuzzy clustering technique. Subsequently, it trains the different ANN using different subsets. Then it determines membership grades of these subsets and combines them via a new ANN to get final results. The whole framework of FC-ANN is illustrated in Fig 1. As typical machine learning framework, FC-ANN incorporates both the training phase and testing phase. The training phase includes the following three major stages:
Stage I: For an arbitrary data set DS, it is firstly divided into training set TR and testing set TS. Then the different training subsets TR1, TR2..., TRk are created from TR with fuzzy clustering module.

Stage II: For each training subset TRi (i = 1, 2..., k), the ANN model, ANNi, (i = 1, 2..., k) is training by the specific learning algorithm to formulate k different base ANN models.

Stage III: In order to reduce the error for every ANNi, we simulate the ANNi using the whole training set TR and get the results. Then we use the membership grades, which were generated by fuzzy clustering module, to combine the results. Subsequently, we train another new ANN using the combined results.

In the testing phase, we directly input the testing set data into the k different ANNi and get outputs. Based on these outputs, the final results can then be achieved by the last fuzzy aggregation module.

The three stages of FC-ANN framework raise three important issues:

1) how to create k different training subsets from the original training dataset TR;
2) how to create different base model ANNi with different training subsets;
3) how to aggregate the different results produced by different base model ANNi

These issues will be addressed by the following sections, respectively.

i) FC Module (Fuzzy Clustering Module)

The aim of fuzzy cluster module is to partition a given set of data into clusters, and it should have the following properties: homogeneity within the clusters, concerning data in same cluster, and heterogeneity between clusters, where data belonging to different clusters should be as different as possible. Through fuzzy clustering module, the training set is clustered into several subsets. Due to the fact that the size and complexity of every training subset is reduced, the efficiency and effectiveness of subsequent ANN module can be improved.

The clustering techniques can be divided into hard clustering techniques and soft clustering technique. Beside partition of training set, we also need to aggregate the results for fuzzy aggregation module. Therefore, we choose one of the popular soft clustering techniques, fuzzy c-means clustering, for fuzzy clustering module.
Ann module

ANN module aims to learn the pattern of every subset. ANN is a biologically inspired form of distributed computation. It is composed of simple processing units, and connections between them. In this study, we will employ classic feed-forward neural networks trained with the back-propagation algorithm to predict intrusion.

A feed-forward neural network has an input layer, an output layer, with one or more hidden layers in between the input and output layer. The ANN functions as follows: each node i in the input layer has a signal xi as network’s input, multiplied by a weight value between the input layer and the hidden layer.

Fuzzy Aggregation Module

The aim of fuzzy aggregation module is to aggregate different ANN’s result and reduce the detection errors as every ANNi in ANN module only learns from the subset TRi. Because the errors are nonlinear, in order to achieve the objective, we use another new ANN to learn the errors.

4) Turing Test Module

In this module the client is provided with some CAPTCHA code which client will input through keyboard, doing this admin will understand that the client is a human not a machine.

5) Question Generation Module

In this module if client fails to perform Login then admin will ask some questions which client has to answer perfectly. The question will be stored by admin at the time of client registration.

Fuzzy cluster technique:

The main thing of fuzzy cluster technique is to dividing wall a known set of data into clusters, and it should have the following properties: homogeneity within the clusters, relating to data in same cluster, and heterogeneity between clusters, where data belonging to different clusters should be as dissimilar as possible. All the way through fuzzy cluster technique, the training set is clustered into several subsets. Due to the fact that the size and complexity of every training subset is abridged, the effectiveness and efficiency of subsequent ANN module can be enhanced.

The fuzzy cluster is composed of the following steps:
Step 1: Initializing Data Sets.

Step 2: manipulative centers vectors

Step 3: Updating Vectors

Step 4: Creating Subset Vectors

**Artificial Neural Network**

ANN component aims to learn the sample of every subset. ANN is an in nature stirred form of distributed estimation. It is collected of simple processing units, and links between them. In this study, we will employ classic feed-forward neural networks skilled with the back-propagation algorithm to imagine intrusion.

A feed-forward neural network has an input layer, an output layer, with one or more concealed layers in between the input and output layer. The ANN functions as follows we will see each node i in the input layer has a signal xi as network’s input, multiplied by a weight value between the input layer and the hidden layer.

**Fig 3 Fuzzy Clustering Module**

**Fuzzy Integration**

The most important target of fuzzy aggregation [5] module is to aggregate dissimilar ANN’s result and reduce the detection errors as every ANNi in ANN module only understand from the
subset TRi. Because the errors are nonlinear, in order to accomplish the purpose, we use another new ANN to study the errors as follows we see stepwise:

Step 1: The total training set TR as data to input the every trained ANNi and obtain the outputs

Step 2: Summarize the input for new ANN

Step 3: Plan the new ANN. We can use Y input as input and use the absolute training set TR’s class label as output to prepare the new ANN.

Fig 4 General Structure

Training Stage

In this stage, we train our model with the aim to prepare it for the test stage. This stage is composed of two steps:

Train the first level: we train the different classifiers of the first level with the training data set, where each feature of the training data set represents an input for the classifier.

Train the second level: a new data set is created from the predictions of the classifiers of the first level. To generate this new Training data set, we associate the selected prediction’s results with the correct label as in the following Table 1. The new training data set is used to train the selected classifier of the second level.
Table 1. The New Training Data Set

Test Stage

In this stage, we test the performance of our model after the achievement of the training stage, where we use the test data set. We process each record of the test data set by the different classifier of the first level. Then, we use the selected prediction outputs of the different classifiers of the first level as an input of the classifier of the second level.

Optimization of Training and Test Time

![Figure 5 Distributed Architecture](image-url)
CONCLUSION

Intrusion detection is an essential component in network security. IDS offers the possible advantages of reducing the manpower needed in monitoring, increasing detection effectiveness, providing data that would otherwise not be obtainable, helping the information security community learn about new vulnerabilities and providing officially authorized proof. In this paper, we survey a new intrusion detection approach, called FC-ANN, based on ANN and fuzzy clustering. Through fuzzy clustering technique, the mixed training set is divided to several homogenous subsets. Thus difficulty of each sub training set is reduced and accordingly the detection performance is improved. The experimental results using the KDD CUP 1999 dataset demonstrates the efficiency of our new approach specially for low-frequent attacks.

REFERENCES


